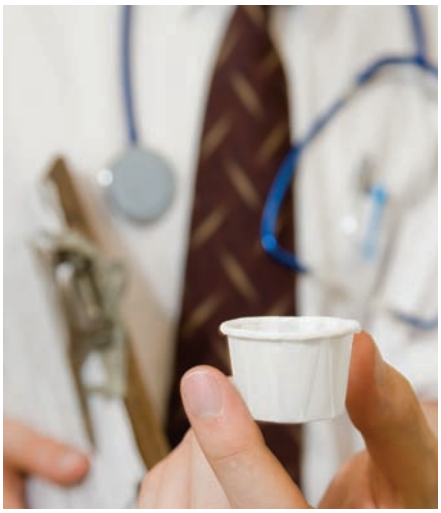

CASE STUDY

How Adaptive Trial Designs Can Increase Efficiency in Psychiatric Drug Development: A Case Study

by JOAN SHEN, MD; SHELDON PRESKORN, MD; VLADIMIR DRAGALIN, PhD; MARY SLOMKOWSKI, PharmD; S. KRISHNA PADMANABHAN, PhD; PARVIN FARDIPOUR, PhD; AMARNATH SHARMA, PhD; and MICHAEL KRAMS, MD, PhD

Dr. Shen is Senior Director, Pfizer Inc., Collegeville, Pennsylvania; Dr. Preskorn is Professor in the Department of Psychiatry and Behavioral Sciences, University of Kansas School of Medicine, Wichita, Kansas; Dr. Dragalin is Senior Vice President, Aptiv Solutions, Raleigh, North Carolina; Dr. Slomkowski is Director, Pfizer Inc., Collegeville, Pennsylvania; Dr. Padmanabhan is Senior Principal Statistician, Pfizer Inc., Collegeville, Pennsylvania; Dr. Fardipour is Senior Director, Pfizer Inc., Collegeville, Pennsylvania; Dr. Sharma is Vice President, Pfizer Inc., Collegeville, Pennsylvania; and Dr. Krams is Vice President, Janssen Pharmaceuticals, Titusville, New Jersey.

Innov Clin Neurosci. 2011;8(7):26–34

FUNDING: No funding was received for the preparation of this article.

FINANCIAL DISCLOSURES: The study used as an example in this article was sponsored by Wyeth Pharmaceuticals. Drs. Shen, Dragalin, Slomkowski, Padmanabhan, Fardipour, Sharma, and Krams were all employees of Wyeth Pharmaceuticals during the time the exampled study was conducted. Dr. Preskorn has been an investigator of Wyeth-funded studies as well as a paid consultant and member of the speakers bureau for Wyeth Pharmaceuticals.

ADDRESS FOR CORRESPONDENCE: Sheldon H. Preskorn, MD, Clinical Research Institute, 8911 E. Orme, Wichita, KS 67207; Email: spreskorn@cri-research.net

KEY WORDS: Clinical trial design, Bayesian statistics, response-adaptive dose allocation, neuropsychopharmacology, schizophrenia

ABSTRACT

This paper uses a recently completed study to illustrate how adaptive trial designs can increase efficiency of psychiatric drug development. The design employed allowed a continuous reassessment of the estimated dose-response such that patients were randomized in a double-blind fashion to one of seven doses of the investigational drug, placebo, or active comparator. The study design also permitted early detection of futility allowing for early study termination. By using the adaptive trial design approach, only 202 patients were needed to make the determination of futility. In contrast, a conventional design would have required enrollment of 450 patients and considerably more time and expense to reach the same conclusion. Adaptive trial designs are important at this time when many pharmaceutical companies are abandoning the development of psychiatric medications because of the inefficiency of conventional approaches.

BACKGROUND

Recently, several major pharmaceutical companies significantly retrenched or stopped doing psychiatric drug development altogether. The principal reasons were two fold: the high cost of such efforts and the high rates of failure. This paper describes how adaptive trial approaches can increase the efficiency of early drug development, using a recently completed study of a novel, investigational antipsychotic drug as an illustration. The focus of the paper is the methodology employed in the

study and the knowledge gained as a result, rather than on the compound *per se*.

Adaptive trial designs allow for the modification of aspects of a study in a real-time, data-driven manner. Such modifications can include 1) stopping early for futility or success, 2) expanding sample size due to greater than expected data variability, or 3) allocating patients preferentially to regimens with a better therapeutic index. However, such potential adaptations must be defined before the trial starts to maintain the trial's validity and integrity.[3–5] Given that the goal is to make adaptations as the study is ongoing, success is dependent on time to information (the earlier, the better) and recruitment speed relative to gathering and analyzing the observations required to adapt.[6,7]

At least two adaptations are frequently considered on an ongoing basis: 1) adapting the sample size due to variability in the observed data and the desired effect size, resulting in a dynamic termination rule (stopping the trial at the earliest time when a predefined level of confidence has been gained that either the dose-response and target dose are sufficiently well identified or that none of the doses achieves the desired treatment effect) and 2) adapting treatment allocation based on accruing response data, so that patients are preferentially allocated to informative doses later in the trial and/or unsafe or underperforming treatment arms are dropped.

The latter adaptation is important because early in drug development the nature of the dose-response relationship is often unknown: Is it a step function, linear, sigmoidal, or curvilinear? The traditional dose-finding studies generally involve only 2 to 3 doses of the investigational drug, whereas many more doses (e.g., 15 or more) can be studied using an adaptive trial approach that utilizes modeling to estimate dose-response curves based on all available data accrued in real time.

This article discusses how these elements were handled in an investigational antipsychotic trial and how use of the adaptive trial design, in contrast to the more conventional approach, permitted more extensive study of the dose-response nature of the drug and resulted in a decision to terminate the study for futility at considerable savings in terms of time and expense.

OBJECTIVE

The objective was to investigate the dose-response of the investigational agent (a selective 5-HT2C agonist) and identify the correct dose(s) for later trials. The hypothesis was that treatment of patients with acute exacerbation of schizophrenia with the investigational drug would produce clinically meaningful improvement on total Positive and Negative Syndrome Scale (PANSS)[9] score without weight gain. References 1, 2, and 8 in the References section provide further rationale for studying the potential efficacy of a 5-HT 2C agonist in schizophrenia.[1,2,8]

MATERIALS AND METHODS

Study design. This was a multicenter, 28-day, double-blind, dose-finding study with randomized response adaptive allocation (based on a computer algorithm) to one of seven doses (50–600mg/day) of the investigational drug (60% of patients), placebo (20% of patients), or active comparator (risperidone 4mg/day) (20% of patients) in hospitalized patients. Data on the primary endpoint, total PANSS score, were collected electronically at Baseline and on Days 7, 14, 21, and 28. The goals of the study were 1) to find the minimally effective dose (MED) that yielded an operationally defined clinically meaningful effect (10 points greater improvement from baseline than placebo on total PANSS by Day 28 with a standard deviation [SD] of 19.4 units) and 2) to estimate the nature of the dose-response curve.

The adaptive design included dose-response modeling, stopping rules, adaptive allocation of subjects, and longitudinal modeling. The design was based on Bayesian modeling in which the unknown model parameters are not expressed in the conventional way as single point estimates (e.g., mean 10 points and SD=19.4), but rather described by a probabilistic distribution. These descriptions based on historical data are called *prior* distributions (i.e., before observing the data) in contrast to *posterior* distributions (i.e., after observing the data). The design had an adaptive treatment allocation and an early termination rule, allowing the trial to be stopped at the earliest point when futility or efficacy was established. Patients were sequentially randomized to placebo, active comparator, or the dose of the investigational drug optimal to characterize the dose-response curve and estimate the MED.

A termination rule based on bounds of posterior probability was used to recommend cessation of recruitment after sufficient information about the dose-response relationship was obtained (i.e., when the following conditions had been satisfied):

- The posterior probability that a particular dose was the MED reached a threshold of 0.6
- The posterior probability that a particular dose was the maximum dose (MaxD) (i.e., the dose that achieves maximum change from baseline on PANSS total score) reached a threshold of 0.6
- The posterior probability that the MaxD achieved a 10-point difference from placebo on total PANSS score was greater than 0.80.

The formulation of the decision rule in terms of posterior probabilities is natural since the posterior distribution tells us what is reasonable to believe given the evidence in data available at a given stage in the trial. For example, the last condition (the third bullet above) can be equivalently formulated in terms of conventional ("frequentist") statistics as the lower bound of the one-sided 80-percent confidence interval for the difference in total

PANSS score of the MaxD and placebo is greater than 10 points. These design control parameters were determined through extensive clinical trial simulations, exploring the design's operating characteristics in a dozen different scenarios.

The maximum sample size for the study was set at 450 patients. By protocol, data from a minimum of 125 evaluable patients randomized into the trial was required before a recommendation of termination for lack of benefit could be endorsed. Stopping for futility was to occur when the posterior probability to achieve a clinically significantly important difference was smaller than 0.01 for all doses.

The algorithm underlying the response adaptive treatment allocation was based on the primary efficacy endpoint alone. Rather than building side effects directly into the model, a Data Monitoring Committee (DMC) reviewed the emerging safety and tolerability information on an ongoing basis to ensure that treatment allocation as recommended by the algorithm was acceptable. The DMC could discontinue allocation to doses based on emerging safety data.

Simulations were used to characterize the behavior of this design for different dose-response curves and conditions. For flat dose responses, median sample size before cessation was 230, with a three-percent false-positive rate. For sigmoid dose-response curves with five-point additional benefit on total PANSS over placebo (i.e., 50% of what had operationally been defined as clinically meaningful benefit) resulted in 39 percent of simulations stopping early: 88 percent of those for lack of benefit and 12 percent due to a false-positive result. If only one dose achieved a clinically meaningful difference, the algorithm allocated patients preferentially to this dose, even with an assumed up-down dose-response curve. In 80 percent of cases, the study would have to fully recruit before a recommendation to stop was made. In scenarios with sigmoid dose-response shapes and benefits of 12.5 points on total PANSS, the system successfully identified the target dose to take into confirmatory trials and stopped early for success in 40 percent of cases.

The study was approved by independent ethics committees and conducted in accordance with guidelines for good clinical practice and the Helsinki Declaration as revised in 1989.

Modeling and statistical analysis. Throughout the trial, a Bayesian normal dynamic linear model (NDLM)[10] was applied to all available PANSS total scores, as assessed by the raters at the clinical site. The NDLM is flexible, requiring no assumptions about monotonicity or shape of the underlying dose-response curve. The prior for the placebo response was assumed to have a mean of zero and a standard deviation of 10. The primary endpoint on active comparator was modeled separately. The change in total PANSS score from baseline was measured every week, with the 28-day score the primary endpoint. A piece-wise linear longitudinal model incorporated the weekly measurements in interim analyses to improve accuracy of estimation of the primary endpoint.

The primary population for final efficacy analysis was the modified intent-to-treat (mITT) population that included all randomly assigned subjects who had taken at least one dose of double-blind test article, had a baseline PANSS total score, and at least one PANSS total score while on therapy. Mixed-effects model based on mITT with observed cases (OC) was used as the primary analysis to assess treatment effects with change in the PANSS total score from baseline as the response variable, treatment (with all the new treatment doses and placebo), visit and treatment by visit interaction as fixed factors, and baseline value as a covariate. An unstructured (UN) covariance was used to model the within-subject repeated measures.

A more detailed discussion of the model-based design, its operating characteristics, and statistical aspects will be presented in a separate statistical methodology paper.

DMC. The core DMC consisted of four sponsor employees—a psychiatrist, a neurologist, a biostatistician, and a programming expert who were not part of the project team or otherwise involved with the trial. The core DMC received weekly estimates of the dose-response relationship, key efficacy and safety data, and updates of the probability of the trial warranting termination for lack of benefit or success and reviewed the computer algorithm's performance.

The full DMC consisted of the core plus five external psychiatrists and internists and was scheduled to meet upon randomization of the 100th, 200th, 300th, and 400th subject.

The DMC's responsibility was to endorse or reject the weekly updates from the adaptive algorithm on the proportion of patients to be allocated to the different treatment arms and to advise study personnel regarding the continuing safety of study subjects and the validity and scientific merit of the trial (i.e., whether to terminate the study early for safety, lack of benefit, or success). For more information on trial committees, see references 6 and 7.

Sites. The study was conducted at 27 sites in the United States.

Subjects. Inclusion criteria consisted of being 18 to 65 years of age, having a *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (*DSM-IV-TR*) diagnosis of schizophrenia (295.10, 295.30, or 295.90), and having a total PANSS between 70 and 120 with positive symptom subscale score >19 and a score >3 on at least two of the following: delusions, conceptual disorganization, hallucinatory behavior, suspiciousness, or unusual thought content at screening and baseline.

TABLE 1. Summary of baseline patient characteristics by treatment group

CHARACTERISTICS	PLACEBO	STUDY DRUG								RISPERIDONE
		50mg	100mg	150mg	200mg	300mg	400mg	600mg	All doses	
Number of subjects (%)	37 (18)	18 (9)	8 (4)	19 (10)	16 (8)	19 (10)	12 (6)	30 (15)	122 (61)	43 (21)
Mean age (range) in years	42 (21–62)	43 (24–57)	41 (34–47)	40 (23–62)	43 (31–57)	41 (23–59)	38 (22–50)	44 (23–57)	41 (22–62)	43 (21–61)
Female gender in %	22	22	13	21	31	16	0	20	19	23
Ethnic origin in % African-American Caucasian	 78 22	 67 28	 50 38	 53 37	 75 25	 79 21	 58 42	 53 43	 62 33	 63 33
Mean weight (range) in kg	90 (57–151)	89 (62–142)	92 (61–147)	84 (51–117)	86 (61–123)	91 (63–146)	89 (63–155)	93 (53–124)	89 (51–155)	92 (42–179)
Duration of current episode at baseline (range) in days	19 (1–53)	33 (7–129)	23 (7–40)	17 (7–41)	19 (1–63)	23 (1–121)	14 (1–35)	21 (3–47)	21 (1–129)	23 (11–89)
Baseline total PANSS (SD)	94.7 (11.2)	95.4 (9.6)	97.3 (14.7)	100.5 (12.8)	93.1 (8.8)	94.4 (8.7)	92.8 (10.3)	95.5 (12.5)	95.6 (8.0)	91.5 (11.0)

PANSS=Positive and Negative Syndrome Scale; SD=standard deviation

Exclusion criteria consisted of having an *DSM-IV-TR* Axis I disorder other than schizophrenia, significant risk of suicide or violent behavior, high or chronic use of benzodiazepines, continuous treatment with anticholinergic drugs, and positive urine drug screen for drugs of abuse.

Subjects gave written informed consent and were hospitalized for the duration of the study.

Outcome measures. Primary efficacy endpoint was change in PANSS total score from baseline to end of the double-blind treatment period (Study Day 28). Total PANSS was also assessed at Days 7, 14, and 21, and a longitudinal model was used to incorporate these early measurements into the interim analyses to improve accuracy of estimation of the primary endpoint. The PANSS was assessed both by the clinical site, which determined enrollment and drove the algorithm, and by a central rater. The latter data was used for parallel but secondary analyses after completion of the study.

Blood samples for pharmacokinetic (PK) analysis were obtained during the study for population PK analysis and to establish retrospectively whether the PK profile for each patient was consistent with the dose assignment as a measure of adherence (results not reported here).

Study medication. All treatments were administered in a blinded fashion using identical-appearing capsules containing 50, 100, or 200mg of the investigational drug or placebo and identical-appearing capsules containing 4mg of risperidone or placebo. Each subject was allocated four bottles and asked to take one capsule from each bottle at the same time each day for 28 days.

Eligible subjects entered a screening period of 4 to 10 days. Initially, subjects who met all entry criteria were equally randomized on Day 0 to placebo, risperidone, or 50, 150, or 300mg of the investigational drug. Allocation to placebo and risperidone was held constant at 20 percent, but the allocation to the different doses of the investigational drug was adaptive in a data-driven fashion while maintaining the blind. After 10 subjects had been randomized to each treatment arm, an interim analysis was conducted and the data reviewed by the DMC. A statistical algorithm used observed data on total PANSS scores between baseline and four weeks of treatment and other associated observations to adapt allocations of subsequent

TABLE 2. Number and percentage of patients who completed or discontinued from the study due to either tolerability/safety issues or lack of efficacy

SUBJECT INFORMATION	PLACEBO	STUDY DRUG							RISPERIDONE
		50mg	100mg	150mg	200mg	300mg	400mg	600mg	All doses	
Number of subjects (%)	37 (18)	18 (9)	8 (4)	19 (10)	16 (8)	19 (10)	12 (6)	30 (15)	122 (61)	43 (21)
Completed (%)	20 (54)	13 (72)	3 (38)	9 (47)	11 (69)	14 (74)	9 (75)	16 (53)	75 (61)	32 (74)
Discontinued due to tolerability/ safety (%)	4 (11)	0 (0)	1 (13)	1 (5)	0	0	0	2 (7)	4 (3)	5 (12)
Discontinued due to lack of efficacy (%)	2 (5)	2 (11)	2 (25)	3 (16)	1 (6)	2 (11)	2 (17)	5 (17)	17 (14)	1 (2)

NOTE: Absolute numbers of patients are followed by percentage within that treatment group. The column between 600mg and risperidone provides a summary of all patients on study drug.

subjects to different doses of investigation drug. At this point, additional dose groups of the investigational drug could be introduced (100, 200, 400, or 600mg).

After the first interim analysis, weekly analyses were conducted using the statistical algorithm, with results provided to the DMC. The randomization scheme could be modified at the end of each analysis until enrollment was complete or a decision for early termination for either success or futility was made.

Concomitant medications. Drugs prohibited during the study included opioid analgesics, investigational drugs, and most other psychotrophic medications. Lorazepam, zaleplon, and zolpidem were permitted as needed for agitation or insomnia.

RESULTS

A total of 280 subjects were screened, and 202 subjects were randomized and treated between December 2007 and May 2008. Thirty-seven subjects were allocated to placebo, 43 to risperidone, and 122 to seven different doses of study drug. Subject demographics and baseline characteristics are shown in Table 1. The mean duration of the acute exacerbation was less than 15 days in 53 percent of the placebo group, 31 percent of the risperidone group, and 23 to 58 percent of the investigational drug groups. Table 2 shows number of patients who completed or withdrew by treatment arm. Study completion rates were 74 percent for risperidone, 61 percent for study drug, and 54 percent for placebo. The number of subjects who withdrew because of tolerability or safety issues (n=13) was similar for placebo and risperidone (12%) and lower for study drug (3%). Unsatisfactory response (i.e., lack of efficacy) was a more common reason for withdrawal for subjects receiving study drug (14%) compared to placebo (5%) and risperidone (2%). Reason for withdrawal was not given for the remaining subjects.

Dose-response curve. The unique feature of the adaptive dose-response design is its ability to provide information in real time, as data emerge, to guide allocation to different doses of investigational drug. Figure 1 provides four snapshots of the estimates of the dose response as they were made available to the DMC over time (Weeks 11, 13, and 18, when the algorithm first recommended stopping the study, and Week 26, when all follow-up data had become available). Figure 2 describes the allocation of subjects to the different treatment arms over time. Note the initial burn-in period of 11 weeks, during which only 50, 150, or 300mg/day of study drug was administered. Figure 3 shows the course of the weekly updated predictive probability of stopping early either for futility or success, as calculated by the algorithm and communicated to the DMC. Note that the predictive probability of futility is high from the start, due to the high placebo response. The DMC recommended stopping the study for futility (posterior probability of futility: 0.97) 18 weeks after first-

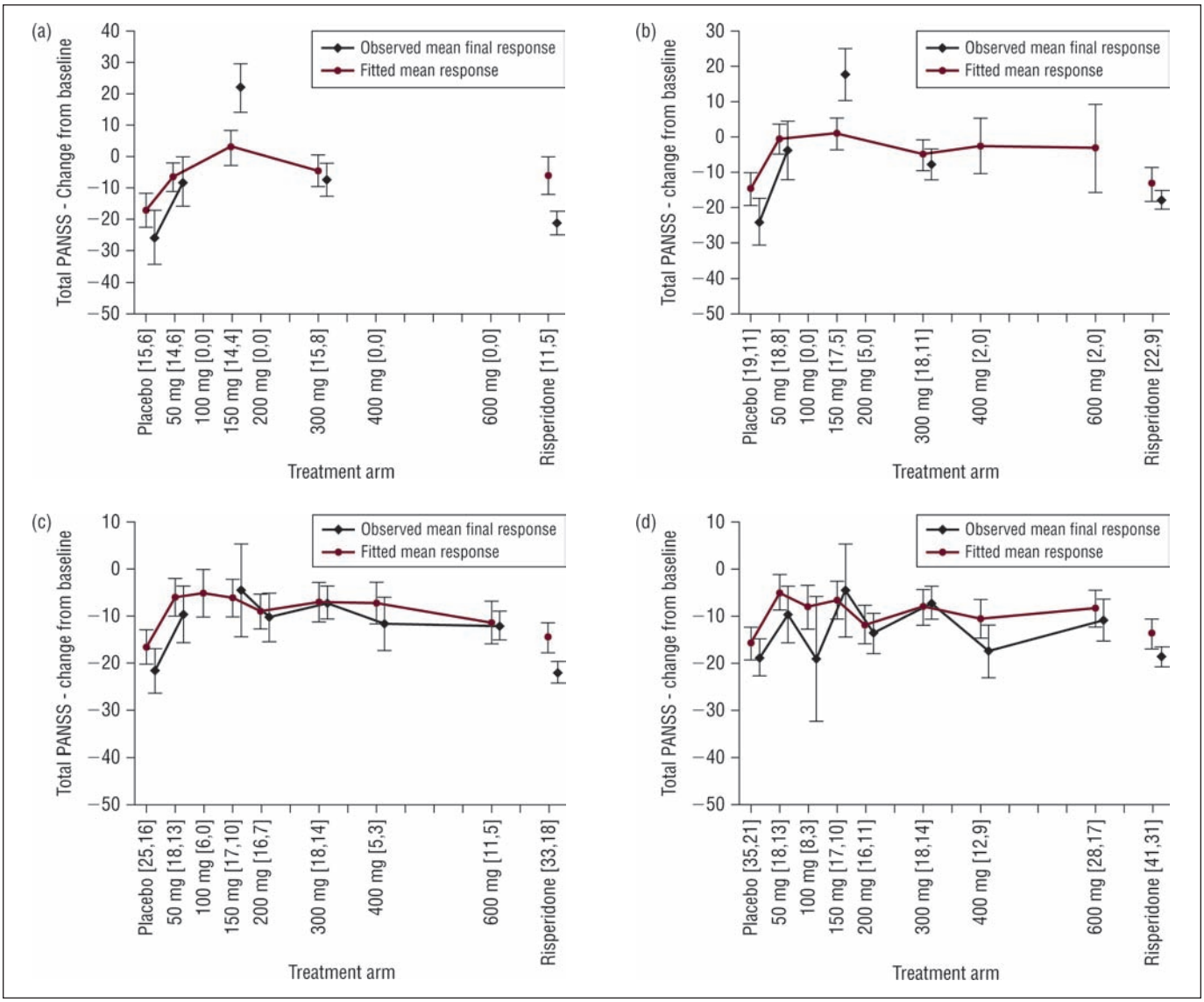

FIGURE 1A–D. Estimates of the dose-response as they were made available to the Data Monitoring Committee over time (Weeks 11, 13, 18, when the algorithm first recommended stopping the study, and Week 26, when all follow-up data had become available). The y-axis describes the change from Baseline to Day 28 on the total Positive and Negative Syndrome Scale (a decrease describes a benefit to the patient). The x-axis lists the different treatment arms, framing the 7 doses of study drug with placebo to the left and risperidone to the right. The first number in [ ] behind the treatment arm indicates the total number of patients with some available data, as used by the model; the second number indicates the number of patients with complete datasets—this is the dataset from which the mean observed (not model based) response was derived. The modeled dose response is shown in black (point estimate for mean response, including 95% confidence interval), and the dose-response curve is interpolated across all doses with available data. The observed (not model-based) data are shown in grey (point estimate for mean response and 95% confidence intervals), and the dose-response curve is interpolated across doses only if final data were available on at least one patient. As the dose-response model excludes the active comparator, the dose-response curve is not interpolated to risperidone. The evolution of information over time is illustrated by the time sequence shown in the figures.

subject-first-visit (FSFV), as only at this timepoint were all the algorithm's stopping criteria met, including having a minimum number of subjects on each treatment arm. Four months (21 weeks) after FSFV, the DMC recommendation was endorsed by the Executive Sponsor Committee and recruitment into the study was stopped.

Clinical outcome. At no stage of the trial did risperidone differentiate from placebo (mean improvement placebo: -18.8, SD±17.9; mean improvement risperidone: -18.7, SD±11.8). The investigational drug did not produce any clinically meaningful or statistically significant effect on any of the efficacy variables at any dose (Table 3). The dose-response curve in Figure 1D (change from baseline to Day 28 on total PANSS) is consistent with a flat-dose response.

Although this study was not intended to examine the relative value of using central versus site raters, central rater data and site rater data were evaluated in a comparable manner and the

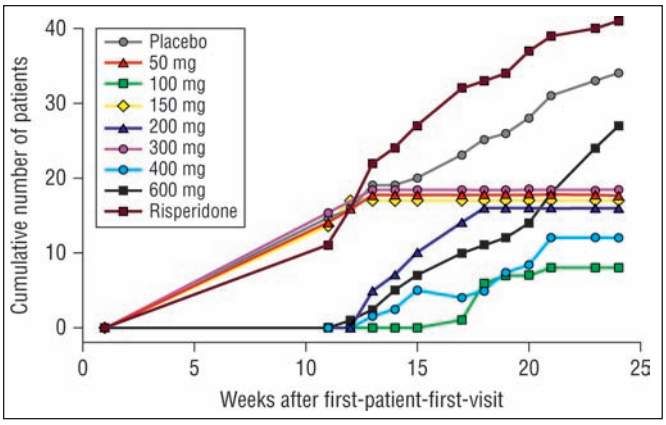

FIGURE 2. Allocation of subjects to the different treatment arms over time. Note the initial burn-in period of 11 weeks, during which only a subset of doses of study drug was available.

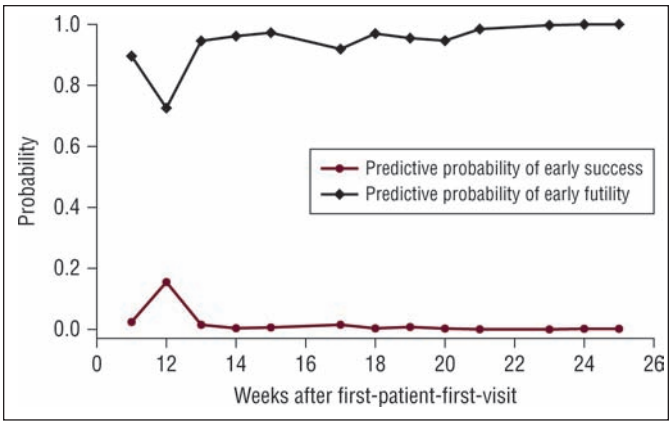

FIGURE 3. Course of the weekly updated predictive probability of stopping early either for futility or success, as calculated by the algorithm and communicated to the Data Monitoring Committee (DMC). Note that the predictive probability of futility is high from the very start, due to the high placebo response. Eighteen weeks after first-subject-first-visit (FSFV), the DMC recommended stopping the study for futility (posterior probability of futility: 0.97), as only at this time point were all stopping criteria of the algorithm met, including having a minimum number of subjects on each treatment arm. Twenty-one weeks after FSFV, the DMC recommendation was endorsed by the Executive Sponsor Committee and recruitment into the study was stopped immediately thereafter.

following findings emerged: The main difference was a trend toward milder PANSS scores, by an average of 10 points, on the central ratings compared to the site ratings. The effect of this shift on recruitment was manually reviewed for all patients who were randomized as a result of their baseline site ratings, and only a small number of patients (n=9, <5%) would not have been randomized had central rather than site ratings been used. Nor would there have been any meaningful difference in outcomes between any of the treatment conditions had central rather than site ratings been used.

Overall, no major safety signal was detected in this study, although there were higher incidences of dizziness and dyspepsia at higher dose levels of the investigational drug. A signal for mild, transient elevation in liver function test (AST/ALT) results was detected in patients on the investigational drug, but no dose-dependent trends were apparent. The profile of the investigational drug for weight and lipid profile was superior to risperidone.

DISCUSSION

In this study, real-time learning about dose-response was deployed in a large, multicenter study of an investigational antipsychotic drug. The adaptive design allowed early determination (i.e., at 21 weeks) of the failed nature of the study (i.e. inability to separate the active control from placebo), which resulted in significant conservation of research resources and fewer patients being exposed unnecessarily to the potential risks inherent in the study of an investigational compound. In comparison, a traditional parallel group design with only three doses, placebo, and active control would have required 400 patients to detect a 10-point mean difference between placebo and one of the investigational treatment groups in PANSS total score with 90-percent power at the 0.05 level of significance assuming a common SD of 19.4. In contrast, the adaptive design employed in this study permitted studying up to seven doses with data-driven assignment of more patients to viable doses and incorporating stopping rules for nonefficacious doses such that the same statistical power could be achieved with a much smaller number of subjects. Of note, time and expense have been significant factors in the recent decisions by some large pharmaceutical companies to abandon psychiatric drug development. Those decisions can result in a lack of progress in combatting psychiatric illnesses, which cause much suffering and loss of productivity and life.

The adaptive design, in particular the use of the NDLM model, allowed the researchers to more efficiently learn about the investigational drug's dose response (i.e., to assess for a potentially nonmonotonic dose-response curve) than would have been possible in a conventional fixed-dose study using three or fewer dosing arms for the investigational drug. A common misunderstanding is that confidence in the findings on a given dose in an adaptive trial design is compromised by the small number of patients on any one of the doses of the investigational drug being tested. However, the modeling approach of

TABLE 3. Mean and standard deviation of the change from baseline to day 28 in total PANSS score for subjects on placebo, the seven doses of study drug, and risperidone

PLACEBO	INVESTIGATIONAL DRUG							RISPERIDONE
	50mg/d	100mg/d	150mg/d	200mg/d	300mg/d	400mg/d	600mg/d	
-18.8±17.9	-9.6±21.8	-19±21.8	-4.5±31.6	-13.6±14.4	-7.2±13	-17.4±16.7	-10.8±18.7	-18.7 ±11.8

the adaptive design uses information from *all* doses in estimating the dose-response curve rather than simply using pair-wise comparisons between study-drug dose arms and placebo.

The proposed adaptive design used modeling only for the dose and efficacy-response relationship. A piece-wise linear longitudinal model was used to estimate the total PANSS score at Week 4 for patients with early withdrawals using their early measurements rather than the last observation carried forward approach. This approach minimizes lack of efficacy being confounded by lack of tolerability or safety issues. Nevertheless, early withdrawals for adverse events can also be directly modeled using the Bayesian approach and used in formal decision rules. However, this approach was not used in this study as it requires more complex models taking into consideration the correlation between the efficacy and safety/tolerability endpoints. The DMC directly viewed the tolerability and safety data as the study progressed to safeguard participants and the integrity of the study.

This use of information from all doses is also germane to the question of whether the failed nature of the study invalidates the dose-response data due to assay insensitivity. A failed study is a function of both placebo and comparator response. A sizable percentage of patients did not respond to placebo but a sufficiently large percentage also did not respond either to the single dose of the comparator (in this case, risperidone) or to any of the doses of the mechanistically novel investigational drug, and there was no evidence of any dose-efficacy relationship. From a drug development perspective, the critical question is whether a decision can be made with confidence as to whether the drug is worth pursuing for this indication based on the totality of the available data. In this case, the decision was to not further pursue the development of this drug for this indication.

Allocation to placebo and risperidone, fixed at 20 percent each (i.e., 40% of all patients), was used to generate reference points. The variability of the data on placebo was nearly twice as great as on risperidone. For this reason, future studies could build into the algorithm the ability to detect greater variance in the placebo group and modify the allocation rules in a predetermined fashion conditioned on emerging data so that a larger number of patients are allocated to placebo and/or active comparator beyond the predetermined 20 percent used in this study to better estimate mean response in patients treated with placebo and/or active comparator.

Running a trial with nine different treatment arms (i.e., risperidone, placebo, and seven doses of the investigational drug) was challenging, both conceptually and logistically. The investigational drug was initially administered in daily doses of 50, 150, and 300mg/day but the design allowed for subsequent doses to be used in a data-driven manner to more fully explore the drug's dose-response curve.

Patient adherence was less of a concern, given that patients were hospitalized and medication was prepared for them by nursing staff. Had this been an outpatient study, blistering study medication and preparing more dose strengths would have been considered. The result of the PK analyses were not presented here in the interest of space and this paper's focus on the adaptive design methodology. Nevertheless, analyses of individual subjects' PK results were consistent with the conclusion that virtually all patients received the doses to which they were allocated.

An unanswered question is: Why did the active comparator fail to separate from placebo? While that is not the focus of this paper, this question may have some impact on the adaptive trial design, especially with regard to the issue of whether a number of the patients enrolled were responsive to hospitalization and the clinical management that they received without medication treatment (i.e., the placebo condition). First, *post-hoc* analyses revealed "study site" as a factor contributing to the placebo response reported here. Recruitment speed differed among sites and was not constant over the course of the study. In fact, there was an initial rapid rate of enrollment (i.e., 202 patients were enrolled in 4 months at 27 US sites) particularly in the initial weeks of the study during the winter months, suggesting that some sites may have had a readily available pool of patients to enroll. As the initial, rapidly enrolling sites completed their number of allotted patients, enrollment slowed. In addition to potentially contributing to the high placebo response rate, this early rapid enrollment could have violated one of the central assumptions of the adaptive designs, exchangeability of patients (i.e., a patient entering the trial in one center early on should be exchangeable with another patient entering the trial in another center at a later stage). The *post-hoc* analysis indicates that assumption was not applicable to this study. In response-adaptive designs, there is an

"optimal" recruitment speed to learn about the research question, and very fast recruitment may be suboptimal. In hindsight, it would have been beneficial to run even more extensive upfront simulations to explore the impact of different scenarios for exchangeability of patients and recruitment speed on the operating characteristics of the design.

Another issue is validity and comparability of ratings at different sites. Site-rated PANSS scores were on average 10 points higher than central-rated scores. However, the variability of observations was comparable, and the difference in scores would not have materially altered the number of subjects enrolled or outcomes for placebo, risperidone, or the different doses of investigational drug. Nevertheless, the algorithm could be easily modified to use central rather than site ratings even though the use of the central raters instead of the site rates would not have altered the outcome of this study.

In our study, only 20 percent of patients received placebo. In major depression trials, it has been argued that ability to establish an efficacy signal decreases with a lower probability of receiving placebo,[11] possibly due to a ceiling effect, through general inflation of response and hence a reduced ability to detect the effect of a pharmacological treatment. This factor could be taken into account when designing the study, while still using the adaptive design methodology.

A failed study by definition involves inability to separate placebo from an active comparator. Hence, the problem could be either with the placebo group or the active comparator group. In this study, only one dose of risperidone (4 mg/day) was used. This dose was chosen for several reasons: 1) it was efficacious compared to placebo in registration trials; 2) it has a lower risk of causing extrapyramidal side effects than higher doses and thus a lower risk of functionally unblinding the patient at the site; 3) it is widely used in the United States and around the world, supporting its clinical efficacy and acceptability. Nevertheless, a higher dose, more than one dose, or even flexible dosing of the active comparator could be built into the adaptive design methodology in future studies. In addition, patients with a history of previous or even current nonresponse to adequate trials of risperidone were not excluded from this trial. Although independent of the adaptive design, this is another factor to consider in future trials. Finally, the use of concomitant sedative medication to control agitation and thus keep patients in the study during the observational period in hospital (Days 1–28) was permitted; however, the degree to which such medication was used was consistent with that in other trials for this indication.

CONCLUSION

The adaptive trial design employed in the study described in this paper was efficient in determining its failed nature and thus limiting the number of patients enrolled and the time and resources utilized. At the same time, this study also highlights the fact that use of such an innovative design methodology alone is not sufficient. There must also be an emphasis on getting the fundamentals of the clinical trial right: choosing the right patient population, optimally assessing the clinical endpoint to enhance the ability to detect a treatment effect, and avoiding confounding factors, such as uncontrolled use of concomitant medications and an urge to recruit quickly.

REFERENCES

1. Di Giovanni G, Di Matteo V, Di Mascio M, Esposito E. Preferential modulation of mesolimbic vs. nigrostriatal dopaminergic function by serotonin (2C/2B) receptor agonists: a combined *in-vivo* electrophysiological and microdialysis study. *Synapse.* 2000;35:53-61.
2. Di Matteo V, Di Giovanni G, Di Mascio M, Esposito E. SB 242084, a selective serotonin-2C receptor antagonist, increases dopaminergic transmission in the mesolimbic system. *Neuropharmacology.* 1999;38:1195–1205.
3. Gallo P, Chuang-Stein C, Dragalin V, et al. Adaptive designs in clinical drug development—an Executive Summary of the PhRMA Working Group. *J Biopharm Stat.* 2006;16:275–283; discussion 285–291, 293–298, 311–312.
4. Krams M, Sharma S, Dragalin V, et al. Adaptive approaches in clinical drug development: opportunities and challenges in design and implementation. *Pharm Med.* 2009;23:139–148.
5. Orloff J, Douglas F, Pinheiro J, et al. The future of drug development: advancing clinical trial design. *Nat Rev Drug Discov.* 2009;8: 949–957.
6. Fardipour P, Littman G, Burns DD, et al. Planning and executing response-adaptive learn-phase clinical trials: 1. The process. *Drug Info J.* 2009a;43:713–723.
7. Fardipour P, Littman G, Burns DD, et al. Planning and executing response-adaptive learn-phase clinical trials: 2. Case studies. *Drug Info J.* 2009b;43:725–734.
8. Millan MJ, Dekeyne A, Gobert A. Serotonin (5-HT)2C receptors tonically inhibit dopamine (DA) and noradrenaline (NA), but not 5-HT, release in the frontal cortex *in vivo*. *Neuropharmacology.* 1998;37:953–955.
9. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:261–276.
10. West M, Harrison J. *Bayesian Forecasting and Dynamic Models, Second Edition.* Springer-Verlag: New York;1997:700.
11. Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol.* 2009;19(1):34–40.